



Overview of Human Variation Resources at NCBI

Selected databases, tools, help documents and educational materials

<https://www.ncbi.nlm.nih.gov/variation>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Introduction

Genome variations and their consequences are of great biological importance to human health, and large volumes of data are becoming available due to rapid progress in sequencing technology and high-throughput analysis. NCBI provides several databases and analysis tools to support research in this field. The NCBI Variation portal (www.ncbi.nlm.nih.gov/variation/) provides a central location from which these databases and tools can be accessed readily.



Databases for Sequence Variations

Several sequence variation databases are available from NCBI. They are dbSNP, dbVar, ClinVar, and dbGaP. Each of these databases are designed to store certain types of data. The table below provides a summary description of these databases and lists key links relevant to each of them.

Variation databases	Description
ClinVar	<p>A public archive of reports on evidence-supported relationships between human variation and observed phenotypes, with submitter-provided assertions of clinical significance.</p> <p>Database www.ncbi.nlm.nih.gov/clinvar Help document www.ncbi.nlm.nih.gov/clinvar/docs/help/ How to submit www.ncbi.nlm.nih.gov/clinvar/docs/submit/ Factsheet ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_Clinvar.pdf Technical support clinvar@ncbi.nlm.nih.gov</p>
dbGaP	<p>A database for genotype and phenotype association datasets from large-scale genome-wide association studies. Personal Identifiable Information, i.e., personal genotype and phenotype data, has restricted access due to privacy and consent concerns. Principal investigators with valid research projects need to apply for access through dbGaP Authorized Access (AA) portal.</p> <p>Database www.ncbi.nlm.nih.gov/gap Help document www.ncbi.nlm.nih.gov/books/NBK5295/ Authorized access dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login How to submit www.ncbi.nlm.nih.gov/gap/docs/submissionguide/ Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_dbGaP.pdf FAQ on dbGaP AA ftp.ncbi.nlm.nih.gov/pub/factsheets/FAQ_dbGaP_Data_Request.pdf Technical support dbgap-help@ncbi.nlm.nih.gov</p>
dbSNP	<p>dbSNP house variation and frequency data from large scale projects including HapMap, 1000Genomes, GO-ESP, ExAC, TOPMED, dbGaP, and HLI to focused studies, such as locus-specific databases (LSDB) and clinical sources. The database is used widely in the fields of personal genomics, medical genetics, and for managing, annotating, and analysis of variation data. dbSNP aggregates genetic variation data from multiple submitters and assigned stable Reference SNP (rs) identifiers that can be used for citation in publication and for integration with other data sources. These rs identifiers are annotated and integrated with other NCBI resources (ClinVar, Gene, RefSeq, PubMed, and BioProject) and disseminated to the scientific community.</p> <p>Database www.ncbi.nlm.nih.gov/snp Help document www.ncbi.nlm.nih.gov/books/NBK3848/ How to submit www.ncbi.nlm.nih.gov/projects/SNP/how_to_submit.html Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_SNP.pdf Technical support snp-admin@ncbi.nlm.nih.gov</p>
dbVar	<p>dbVar is a database of genomic structural variation – insertions, deletions, duplications, inversions, mobile element insertions, translocations, and complex chromosomal rearrangements. It hosts data from 1000Genomes, ExAC, and studies from ClinGen, OMIM, GeneReviews, and ClinVar that defines the clinical relevance of genes and variants for use in precision medicine. Popular studies are listed at www.ncbi.nlm.nih.gov/dbvar/content/human_hub/.</p> <p>Database www.ncbi.nlm.nih.gov/dbvar Help document www.ncbi.nlm.nih.gov/dbvar/content/help/ How to submit www.ncbi.nlm.nih.gov/dbvar/content/submit/ Factsheet ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_dbVar.pdf Technical support dbvar@ncbi.nlm.nih.gov</p>

Tools for Variation Analysis

NCBI also provides several tools for interactive examination of variations under a genomic context, and for analysis of experimental or publicly available variation data within the context of annotated features on the genome. The Remap service, similar to lift-over (UCSC) but with more reliable mapping, is also available for converting genomic coordinates between different assemblies. More details and links are in the table below.

Tools	Description
1000 Genomes Browser	This browser allows the interactive examination of genotype data from the 1000 Genomes project in the context of NCBI's GRCh37p13 assembly and annotation. Genotype data for a region can also be downloaded for use elsewhere. Tool www.ncbi.nlm.nih.gov/variation/tools/1000genomes/ Help www.ncbi.nlm.nih.gov/variation/tools/1000genomes/help/ Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_1000genomes_browser.pdf
Variation viewer	This browser provides a summary of variations mapped to a user-defined region on the genome. Users can navigate by gene or exon, filter results by allele frequencies from various large-scaled studies, or by other criteria, and link to related information in medical genetics resources such as ClinVar, MedGen and GTR. Variants displayed can be downloaded from specific track of interest. Tool www.ncbi.nlm.nih.gov/variation/view/ Help www.ncbi.nlm.nih.gov/variation/view/help/ Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_Variation_Viewer.pdf
Remap service	Use this tool to remap variations and other annotations between different assemblies or assembly versions. Use the "Clinical Remap" option to remap between assemblies and RefseqGene/LRG (Locus Reference Genomic) records. Tool www.ncbi.nlm.nih.gov/genome/tools/remap Help www.ncbi.nlm.nih.gov/genome/tools/remap/docs/whatis Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_Remap.pdf
Phenotype Genotype Integrator (PheGenI)	Use PheGenI to search for summary information derived from GWAS studies reported from dbGaP and NHGRI's GWAS catalog through controlled phenotypic query terms, genomic locations, or features such as rsIDs or genes. Tool www.ncbi.nlm.nih.gov/gap/phegeni Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_PheGenI.pdf
Variation Services (SPDI)	Use this new service to disambiguate variants expressed in different formats to find the canonical variant expressed in the new SPDI format, parse HGVS or VCF input into the new SPDI format, or convert a given variant into left-shift VCF or right-shift HGVS format. Tool api.ncbi.nlm.nih.gov/variation/v0/ Demo www.ncbi.nlm.nih.gov/variation/services/demo/ Blog ncbiinsights.ncbi.nlm.nih.gov/2017/02/09/ Video www.youtube.com/watch?v=gpSI16e59ig Factsheet ftp.ncbi.nih.gov/pub/factsheets/Factsheet_VariationService_SPDI.pdf

Portal for submission of variation data

Variation data housed in ClinVar, dbSNP, and dbVar are based on data submitted by various research groups. Users interested in submitting data to those resources can log into the Variation Submission Portal to register and submit.

- dbSNP/dbVar Submission submit.ncbi.nlm.nih.gov/subs/variation/
- CLinVar Submission submit.ncbi.nlm.nih.gov/clinvar/
- dbSNP/dbVar Help www.ncbi.nlm.nih.gov/variation/submit/
- Clinvar Help www.ncbi.nlm.nih.gov/clinvar/docs/submit/

Introduction to Key Variation Resources

Here are a list of single-page introduction of key variation resources, packaged based on user personas.

- For Genomic Researchers <https://go.usa.gov/xGgMe>
- For Translational Researchers <https://go.usa.gov/xGgMS>
- For Bioinformatic Specialists <https://go.usa.gov/xGgMh>
- For Clinical Labs <https://go.usa.gov/xGgeC>
- For Clinicians <https://go.usa.gov/xGger>